December 1945

BIOMETRICS BULLETIN

Vol. 1 No. 6

THE BIOMETRICS SECTION, AMERICAN STATISTICAL ASSOCIATION

SOME APPLICATIONS OF STATISTICAL METHODS TO FISHERY PROBLEMS*

WILLIAM E. RICKER
Indiana University

Graphical and Numerical Summaries of Data. In its applications to fishery biology, statistics may be said to include the art of presenting data in compact and comprehensible form, and the technique of testing the significance of observed differences or of possible causal relationships. In the first category may be placed the use of graphical representation to show clearly the frequency distribution of length and of other characters in samples taken from fish populations. In an early use of this method, C. G. J. Petersen interpreted successive modes on a length-frequency graph as indicating successive *age-classes* of the fish in question—a procedure which is still valuable, especially for younger age-groups. However, early applications of this method to certain fishes led to unhappy results, owing to the unsuspected existence of marked variations in the abundance or "strength" of different year-classes of fish. If a given year proves exceptionally favorable for reproduction or survival of young, its progeny will be represented more abundantly than are those of adjacent years on a length-frequency graph, and consequently modes are produced which cannot be interpreted as representing *successive* age-groups. These relationships were clarified when the technique came into general use of determining the age of individual fish from marks on the scales, otoliths or other hard parts. It was found that age-frequency histograms from representative samples usually show at least a moderate degree of fluctuation in the strength of successive age-classes, and at times this variation is very marked. A famous example is of the 1904 year-class of herring in Norwegian waters, which was shown by Johan Hjort (15) to be the most numerous age-group present in catches from 1907 through 1918.

Length-frequency graphs can also be used, when necessary, to estimate the rate of growth of fish, since the distance between the position of a mode in one year, and the mode for the same year-class in the following year, is an estimate of the mean increase in length of the fish concerned. Because of the tendency for adjacent ages to crowd together and obliterate the modes in later years of life, it may be necessary to wait for and use a "dominant" year-class, in order to follow growth to any advanced age by this method. In such a situation it is very advantageous to plot each year's deviation from a long-term average, instead of ordinary length-frequency polygons, as may be seen in a paper by Frances Clark (6). Another aspect of length-frequency graphs has been treated by Buchanan-Wollaston and Hodgson (5). These authors show that, within a single age-class of young fish, irregularities in the shape of the fre-

*Contribution No. 347 from the Department of Zoology, Indiana University.

quency distribution sometimes have continuity in successive samples, and can be interpreted as the result of the merging of discrete population units, corresponding possibly to broods hatched at different times of year, or in different places. It is further observed (2) that in order to trace these units over a period of time, with maximum efficiency, it is desirable to use a grouping interval which increases in breadth as the standard deviation in length of the age-class increases.

Graphical presentation has a danger, in that some of the relationships or differences which are suggested may not stand up under critical analysis; and if such analysis is not made, misleading conclusions may be drawn. An illustration is provided by the work of Schnakenbeck (31), who used differences in the shape of the frequency distributions of vertebral counts to define different races of herring in the North Sea. However the observed range of vertebral numbers is quite small, and Buchanan-Wollaston (3) was able to show that the differences in shape could all be accounted for by considering each series of vertebra counts as a highly-grouped normal frequency distribution, the shape depending on the position of the mean in relation to the center of the grouping interval.

Nongraphical summarizing of bodies of data includes, for example, the use of the arithmetic mean, mode and median as measures of average character, and the use of the range, quartiles, deciles, average deviation and standard deviation as measures of variation. All of these have found uses in fishery work, with the emphasis on the arithmetic mean and the standard deviation. However, as a purely descriptive statistic the standard deviation has a serious limitation, since it does not specify the shape of the frequency distribution obtained; and in fishery work, distributions showing considerable deviation from the normal are frequently met. Another widely used arithmetical method of making data more informative is that of smoothing by moving averages.

Even when data have been summarized by these or other numerical methods, graphical illustration may be employed to present the summaries forcefully to the eye. Hubbs and Perlmutter (17) have proposed a system for presenting the mean, range, standard deviation and twice the standard error of the mean all on one graph, illustrating it with a "cline" of vertebral numbers in the anchovy. Others have plotted the mean, and limits corresponding to plus and minus one, two, or three standard deviations. The last-named limits are used by Rich (22) in a "control chart" illustrating production of the Columbia River salmon fishery.

Significance of Differences of Means. The last two examples above illustrate the fact that it is impossible to draw a hard and fast line between the use of statistics for purposes of representation, and their use for testing the significance of differences. Thus Hubbs and Perlmutter's graph, while primarily pictorial, includes the standard error of the mean; and this can be used only for estimating the significance of differences between means, for it is not a descriptive statistic and tells nothing about the population from which the sample was drawn. Similarly Rich plots the "control" level $\pm\sigma$ in order to have a basis for deciding whether the catch in recent years has fallen outside of the previously normal range of variation.

Comparisons of means based on large sample theory, using the normal probability integral, date well back; and in the application of the small sample procedure introduced by R. A. Fisher, fishery workers have probably been as progressive as most biologists. One use of these techniques has been to establish the reality of differences in various physical features of fish of different populations, sexes, broods, ages, etc. Thus Rounsefell (27) and Tester (34) have shown that Pacific herring can be divided into numerous essentially discrete and persistent populations on the basis of the mean number of vertebrae in each, and that this mean number increases from south to north, in a general way. They also show (29, 35) that there are significant differences in number of vertebrae between herring spawned in different years. Both these observations are related to the fact that low temperature during early development makes for a larger number of vertebrae. Hile (14) found that significant and rather large differences exist in the relative size of the head, eyes, fins, and in the number of fin rays, and gill-rakers, in different year-classes of ciscoes in one lake—which discovery cast grave doubt on the feasibility of using such characters to distinguish different *races* of the same

fish, as had previously been done on an extensive scale. Similarly, Mottley (18) questions much of the current use of scale-counts as a systematic character in trout, following a demonstration that number of scales is significantly and importantly affected by temperature during early development. In addition to their possible taxonomic value, some of these studies have great practical importance, in that it is necessary to know whether all the fish in a given fishery form one homogeneous population, or whether they consist of discrete units which, quite possibly, may require individual regulations.

As regards small sample procedure for testing average differences, an application of Foerster and Ricker (11) may be cited. In assessing the effect of a method of management (removal of predacious fishes) on the survival of young salmon, they were able to demonstrate significant improvement following three years of predator control. The comparison of standard deviations or variabilities is also common enough, though much less common than the comparison of means. As an example the work of Dawes (9) can be mentioned, who found differences in the variability in the rate of growth of plaice held at two different locations on the English coast.

Correlation and Regression. The search for associations and causal connections between observed phenomena has often been conducted by computing a correlation or regression coefficient, and testing its significance. In this way Rounsefell (28) has computed a linear relationship between the logarithm of the area of bodies of fresh water and the logarithm of the weight of or yield from their fish populations per acre; though naturally the standard deviation from the regression line ("standard error of estimate") is quite large. Foerster (10) showed that with an increase in population numbers the size of the young salmon in a lake decreased, and the writer (24) found an inverse correlation between the number of salmon present in the same lake and the abundance of their food organisms.

Another use of simple correlation has been in connection with the growth in length of fishes at different ages. Hubbs and Cooper (16) found positive correlations between first and second-year growth in certain centrarchids, while in later years negative correlations may usually be obtained. The net result is that the standard deviation in length of any brood at first increases, and in later years decreases (25). The second half of this process has been treated by various authors under the title "growth compensation." With some species the initial increase in variability is apparently completed by the end of the first growing season, being facilitated by the fact that fry are hatched at various times through the season, hence have been growing for different lengths of time when growth is suspended at the onset of cold weather.

Somewhat more elaborate uses of correlation and regression have been made in at least two other investigations. Dawes (9) computes correlation coefficients and regression equations for weight increase observed in experiments with plaice, as compared with food consumed, and he also attempts to test the linearity of the regressions observed, by means of "correlation ratios"—a method now discarded. Davidson and coworkers (8) have made an interesting application of Sewall Wright's "path coefficients." By this method quantitative estimates were obtained of the direct and indirect influence of each of two correlated factors, namely stream level on two successive days, upon the migration of salmon on the second day; and also of the influence of the rainfall of the two

Application for entry as second-class matter is pending. The Biometrics Bulletin is published six times a year—in February, April, June, August, October and December—by the American Statistical Association for its Biometrics Section. *Editorial Office*: 1603 K Street, N. W., Washington 6, D. C.

Membership dues in the Biometrics Section and the American Statistical Association combined are $6.00 per year, including the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, the BIOMETRICS BULLETIN, and the ASA BULLETIN; for Associate Members of the Section dues are $2.00 per year, which includes the BIOMETRICS BULLETIN. Single copies are 60 cents and annual subscriptions, $3.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N. W., Washington 6, D. C.

days upon the stream level. Probably numerous additional uses for this procedure will be found.

The χ^2 Distribution. The χ^2-test has also come in for its fair share of use by fishery workers. An interesting early application is that of Baranov (1), who was interested in determining the probable degree of divergence between a fish population and a sample taken from it, in respect to frequency of representation of different length classes. He came to the general conclusion that samples of the order of 1000 fish were desirable in order to achieve a reasonable level of accuracy. Chi-square was used by Buchanan-Wollaston (3) to test the frequency distributions of herring vertebral numbers for departures from normality, in the study mentioned earlier. These distributions, incidentally, were so highly grouped that Sheppard's adjustment did not sufficiently correct for the grouping; in order to obtain an accurate estimate of the variability of the population, Buchanan-Wollaston, under the guidance of R. A. Fisher and his Rothamsted colleagues, made a direct attack using the method of maximum likelihood.

An interesting problem to which χ^2 has been applied is that of estimating the randomness of the dispersion of tagged fish. Schumacher and Eschmeyer (32) in this way compared the movements of tagged bass of three species in Norris Lake, Tennessee; and they used the same test to discover whether any significant mortality occurred because of the presence of the tag. Another aspect of the dispersion of tagged fish is treated by Hart (13), who found that the pattern of recoveries of tagged pilchards appeared to change in three respects, according to whether the recoveries were made in the season immediately after tagging, or in later seasons. Probabilities computed for each of the three respects were below ordinarily accepted levels of significance, but when the three were combined a significant change in behavior was confirmed.

Another use of χ^2 was made by Robertson (26), to test various possible causes of *Lee's phenomenon* in samples of the sprat. In fishes which have distinct annual marks on their scales, a calculation of size at all earlier ages can often be made from a scale of any given age, the relation between scale size and fish size usually being fairly simple, and in any event susceptible to observation. It is often found, however, that the *calculated* average size of a given year-class at any age is the smaller, the older are the fish from whose scales it is determined. This effect is named for its discoverer, Rosa Lee, and has been discussed at length by her and by a number of other authors. The two more important possible explanations involve either (1) a greater mortality rate throughout life among faster-growing fish in the population; or (2) selective sampling of the population. In the work cited, by comparing the scale-body ratios for (a) fish of the same year-class in different seasons, (b) fish of the same age-class in different seasons, and (c) fish of different age-classes in the same season, Robertson showed that the "phenomenon" was solely the result of selective sampling. This result does not necessarily apply to all species of fish, perhaps not even to all populations of sprats.

Confidence Limits for Binomial and Poisson Distributions An investigator who gives serious attention to the numerous types of systematic error which may enter into his procedures for sampling fish populations may at times tend to forget that even his best efforts will not make the result anything better than a good *sample*, subject to sampling error. Baranov's examination of this point has already been mentioned. The present writer has found it salutary to make frequent reference to limits of confidence for the binomial and Poisson frequency distributions (7, 12, 23). Almost any character can be treated in this way: e.g. in a sample, the fish of a given species, or size, or age, or sex, or degree of parasitization, can be selected for special consideration, and the ratio of their number to that of the remaining fish, or any suitable part of them, can be considered as an estimate of the corresponding ratio in the population. Then if sampling was random, the appropriate binomial or Poisson fiducial limits will give an estimate of *sampling* variability without further ado. The Poisson limits should be used when fish having the selected character are only a small fraction of the total sample—about 5 percent or less. However, because the Poisson limits are easier to use than binomial limits, and will never give too narrow fiducial limits, they can be employed as a rough

approximation even where this condition is not met.

Even the fact that sampling may actually not be random is no obstacle to the use of binomial or Poisson limits of confidence, for they will usually provide a *minimal* estimate of sampling error—that part of it which cannot be exorcised by any conceivable refinement of technique, though of course it can be decreased by taking a larger sample. One valuable service which the limits may at times perform is to suggest that a possible improvement in sampling procedure will not be worthwhile, and that the extra effort could more profitably be spent in taking a larger sample by the old methods. Equally they may indicate that the sample has become as large as or larger than is necessary for the purpose in mind, or considering the known or probable magnitude of systematic errors. These limits do not fill the same place as direct estimates of the sampling error made from replications, and by comparing them with the latter it is possible to see whether the actual error exceeds what might be due to random sampling.

The occasions on which limits of confidence for these two distributions may profitably be consulted in fishery investigations are literally innumerable, though they need rarely appear in published reports. Unless the needs of other types of investigation differ radically, they would be among the most frequently used tables in a statistical handbook, once investigators generally became familiar with them.

Analysis of Variance and Covariance. The impact of many of the newer statistical methods has still to make itself completely felt in fishery research. The first use of the analysis of variance in this field that has come to my attention is by Buchanan-Wollaston (3); and it was not a simple one, since adjustments for nonorthogonality are introduced. By this means the fact was confirmed that the mean number of vertebrae in North Sea herring varied significantly in respect to area of capture and, in the samples which he had, in respect to age or state of maturity; though the latter may have been a result of year-class variability. In Europe Buchanan-Wollaston is without question the principal fishery biologist to use and advocate the use of statistical methods above the elementary level, and he has very laudably endeavored to convince his colleagues that the new formulae should have no terrors (4). On this continent some of Fisher's methods have found a forthright exponent in C. McC. Mottley, who has communicated his enthusiasm to a promising group of students and associates. This school has applied the methods of analysis of variance and covariance to various problems; for example, that of examining the relation between the abundance and the size of fish in a lake (19), or of testing the validity of the age-old belief—firmly held by a large proportion of anglers—that fish bite better in the dark of the moon (21).

Two attempts to use the Latin square and similar experimental designs, such as are commonly employed in agricultural trials, are known to the writer, though in neither case have reports yet appeared. These concern testing the effects of different cropping methods upon the yield of clam beds, and testing the power of soils added to the bottom marl of barren lakes to promote the growth of aquatic vegetation. Such designs could also advantageously be used in production experiments with pond fish, or even in testing management procedures on a series of native bodies of water, as Mottley (20) has in fact already proposed.

Statistical Estimation. The development of a new "statistic" or formula appropriate to a special need, and of a type not mentioned in available literature, will as a rule be beyond the ability of the average biologist, even though he is familiar with many statistical procedures. In such a situation it is necessary to call for the assistance of a colleague trained to a higher level of mathematical skill. The most interesting example of this sort in fishery work concerns the estimation of populations from marking experiments. A method long used to estimate fish populations consists of marking and releasing a sample of fish into a body of water so that thorough mixing of marked and unmarked is assured, after which the number of marked ones present is observed in a new sample. From the number of marked ones released (B), and the number of marked ones (C) found in the second sample (A), the population (N) is estimated by a simple proportion: $N = AB/C$. When it is impossible to catch a large number of fish for marking in a short space of time, it is economical to con-

duct the experiment so that both marking and recoveries extend concurrently over a period of a few weeks, as was done independently in the early 1930's by David Thompson and Chancey Juday. The problem of finding the best pooled population estimate from all the days of such an experiment was turned over to Zoe Schnabel (30), who used the method of maximum likelihood to deduce that this best estimate would be obtained by solving the equation:

$$\Sigma\left[\frac{B(A-C)}{N}\left(1+\frac{B}{N}+\frac{B^2}{N^2}+\cdots\right)\right]=\Sigma[C],$$

where the symbols A and C are as above, for each day's catch, and B is the total number of marked fish at large on the same day. When B/N and hence C/A are small, this reduces to the approximation:

$$N = \frac{\Sigma(AB)}{\Sigma(C)} \,,$$

which is the principal fromula in actual use.

Schumacher and Eschmeyer (33) attacked the same problem by a different method, that of minimizing squares of residuals, and using a different kind of approximation—namely, the assumption that the value of each day's information is independent of the number of marked fish at large in the population on that day. This assumption yields the formula:

$$N = \frac{\Sigma(AB^2)}{\Sigma(BC)} \,.$$

From an exchange of letters with Dr. Schumacher it appears that the efficiency of this expression is at a maximum when B/N is equal to 0.5, whereas Schnabel's second, or approximate, formula becomes most efficient as $(B/N)\rightarrow 0$, and the two formulae are of equal efficiency when $B/N = 0.25$. Consequently Schnabel's form will ordinarily be best, since the value of B/N rises gradually from a very small initial magnitude, and, except on quite small bodies of water, will not often exceed 0.25 even when the experiment comes to an end. Of course Schnabel's long formula, carried to several terms, can always be used if the best possible estimate is desired; but the labor of the computation will rarely be warranted, considering the magnitude of the sampling and probable systematic errors in such experiments.

Schumacher and Eschmeyer's work however has a merit which Schnabel's lacks, i.e., a formula for computing the standard error of the estimate obtained. The writer (25) has assumed, in situations where Schnabel's approximate formula could be used, that fiducial limits for the Poisson distribution applied to $\Sigma(C)$ would give at least a fairly good idea of the minimum of variability ascribable to random sampling; but an estimate of error obtained directly from the data themselves, for both the general and the special case, is to be desired.

LITERATURE CITED

1. Baranov, F. I. On the question of the biological basis of fisheries. Nauchnyi issledovatelskii iktiologisheskii Institut, Izvestiia, 1 (1), (1916): 81-128. 1918.
2. Buchanan-Wollaston, H. J. Some remarks on the graduation of measurement data. Jour. Counseil Expl. Mer 6: 47-63. 1931.
3. Buchanan-Wollaston, H. J. Some modern statistical methods: their application to the solution of herring race problems. Jour. Conseil Expl. Mer 8: 7-47. 1933.
4. Buchanan-Wollaston, H. J. The philosophic basis of statistical analysis. Jour. Conseil Expl. Mer 10: 249-263; 11: 7-26. 1935-36.
5. Buchanan-Wollaston, H. J. and William C. Hodgson. A new method of treating frequency curves in fishery statistics, with some results. Jour. Conseil Expl. Mer 4. 207-225. 1929.
6. Clark, Frances N. Interseasonal and intraseasonal changes in size of the California sardine (**Sardinops caerulea.**) Calif. Div. Fish and Game, Fish Bulletin No. 47, 28 pp. 1936.
7. Clopper, C. J., and E. S. Pearson. The use of confidence or fiducial limits applied to the case of the binomial. Biometrika 26: 404-413. 1934.
8. Davidson, F. A., Elizabeth Vaughan, S. J. Hutchinson and A. L. Pritchard Factors influencing the upstream migration of the pink salmon (Oncorhynchus gorbuscha) Ecology 24: 149-168. 1943.
9. Dawes, Ben. A statistical study of growth and maintenance in the plaice. Jour. Marine Biol. Assn., N. S. 17: 949-975. 1931.
10. Foerster, R. E. The relation of lake population density to size of young sockeye salmon (Oncorhyncuus nerka). Jour. Fish. Res. Board Canada 6: 267-280. 1944.
11. Foerster, R. E., and W. E. Ricker. The effect of reduction of predaceous fish on survival of young sockeye salmon at Cultus Lake. Jour. Fish. Res. Board Canada 5: 315-336. 1941.
12. Garwood, F. Fiducial limits for the Poisson distribution. Biometrika 28: 437-442. 1936.
13. Hart, John Lawson. Tagging experiments on British Columbia pilchards. Jour. Fish. Res. Board Canada 6: 164-182. 1943.
14. Hile, Ralph. Morphometry of the cisco, **Leucichthys artedi** (LeSueur), in the lakes of the

Northeastern Highlands, Wisconsin. Int. Revue Hydrobiol. 36: 57-130. 1937.
15. Hjort, Johan. Fluctuations in the year classes of important food fishes. Jour. Conseil Expl. Mer 1: 5-38. 1926.
16. Hubbs, Carl L., and Gerald P. Cooper. Age and growth of the long-eared and the green sunfishes in Michigan. Papers Michigan Acad. Sci. 20 (1934): 669-696. 1935.
17. Hubbs, Carl L., and Alfred Perlmutter. Biometric comparison of several samples, with particular reference to racial investigations. Amer. Naturalist 76: 582-592. 1942.
18. Mottley, C. McC. The effect of temperature during development on the number of scales in the Kamloops trout, **Salmo kamloops Jordan.** Contro Can. Biol. and Fish. 8: 255-263. 1934.
19. Mottley, C. McC. Effect of increasing the stock in a lake on the size and condition of rainbow trout. Trans. Amer. Fish. Soc. 70 (1940): 414-420. 1941.
20. Mottley, C. Mc. Experimental designs for developing and testing a stocking policy. Trans. 7th N. Amer. Wildlife Conf., pp. 224-238. 1942.
21. Mottley, C. McC., and Daniel R. Embody. The effect of the full mooh on trout fishing Jour. Amer. Stat. Assn. 37: 41-47. 1942.
22. Rich, Willis H. An application of the control chart method to the analysis of fisheries data. Science 97: 269-270. 1943.
23. Ricker, William E. The concept of confidence or fiducial limits applied to the Poisson frequency distribution. Jour. Amer. Stat. Assn. 32: 349-356. 1937.
24. Ricker, William E. The food and the food supply of sockeye salmon **(Oncorhynchus** Nerka Walbaum) in Cultus Lake, British Columbia. Jour. Biol. Board Canada 3: 450-468. 1937.
25. Ricker, William E. Creel census, population estimates and rate of exploitation of game fish in Shoe Lake, Indiana. Investigations of Indiana Lakes and Streams 2: 215-253 1942.
26. Robertson, J. Armitage. The occurrence of Lee's phenomenon in the sprat, and the size-relation between fish and scale. Jour. Conseil Expl. Mr 11: 219-228. 1936.
27. Rounsefell, George A. Contribution to the biology of the Pacific herring, **Clupea pallasii,** and the condition of the fishery in Alaska. Bull. U. S. Bur. Fish. 45: 227-320. 1930.
28. Rounsefell. George A. Fish production in lakes as a guide for estimating production in proposed reservoirs. U. S. Fish and Wildlife Service, 14 pp., mimeographed. 1945.
29. Rounsefell, G. A., and E. H. Dahlgren. Fluctuations in the supply of herring, **Clupea pallasii,** in Prince William Sound, Alaska. Bull. U. S. Bur. Fish. 47 (9): 263-291. 1932
30. Schnabel, Zoe E. The estimation of the total fish population of a lake. Amer. Math. Monthly 45: 348-352. 1938.
31. Schnakenbeck, W. Zum Rassenproblem bei den Fischen. Zeitschrift fur Morphol. u. Okologie der Tiere 21 (3/4). 1931.
32. Schumacher, F. X., and R. W. Eschmeyer. The recapture and distribution of tagged bass in Norris Reservoir, Tennessee. Jour. Tennessee Acad. Sci. 17: 253-268. 1942.
33. Schumacher, F. X., and R. W. Eschmeyer. The estimate of fish population in lakes or ponds. Jour. Tennessee Acad. Sci. 18: 228-249. 1943.
34. Tester, Albert L. Populations of herring **(Clupea pallasii)** in the coastal waters of British Columbia. Jour. Fish. Res. Board Canada 3: 108.144. 1937.
35. Tester, Albert L. Variation in the mean vertebral count of herring **(Clupea pallasii)** with water temperature. Jour. Conseil Expl. Mer 13: 71-75. 1938.

INDIVIDUAL COMPARISONS BY RANKING METHODS

FRANK WILCOXON
American Cyanamid Co.

The comparison of two treatments generally falls into one of the following two categories: (a) we may have a number of replications for each of the two treatments, which are unpaired, or (b) we may have a number of paired comparisons leading to a series of differences, some of which may be positive and some negative. The appropriate methods for testing the significance of the differences of the means in these two cases are described in most of the textbooks on statistical methods.

The object of the present paper is to indicate the possibility of using ranking methods, that is, methods in which scores 1, 2, 3, . . . n are substituted for the actual numerical data, in order to obtain a rapid approximate idea of the significance of the differences in experiments of this kind.

Unpaired Experiments. The following table gives the results of fly spray tests on two preparations in terms of percentage mortality. Eight replications were run on each preparation.

Sample A		Sample B	
Percent kill	Rank	Percent kill	Rank
68	12.5	60	4
68	12.5	67	10
59	3	61	5
72	15	62	6
64	8	67	10
67	10	63	7
70	14	56	1
74	16	58	2
Total 542	91	494	45

Rank numbers have been assigned to the results in order of magnitude. Where the mortality is the same in two or more tests, those tests are assigned the mean rank value. The sum of the ranks for B is 45 while for A the sum is 91. Reference to Table I shows that the probability of a total as low as 45 or lower, lies between 0.0104 and 0.021. The analysis of variance applied to these results gives an F value of 7.72, while 4.60 and 8.86 correspond to probabilities of 0.05 and 0.01 respectively.

Paired Comparisons. An example of this type of experiment is given by Fisher (2, section 17). The experimental figures were the differences in height between cross- and self-fertilized corn plants of the same pair. There were 15 such differences as follows: 6, 8, 14, 16, 23, 24, 28, 29, 41, —48, 49, 56, 60, —67, 75. If we substitute rank numbers for these differences, we arrive at the series 1, 2, 3, 4, 5, 6, 7, 8, 9, —10, 11, 12, 13, —14, 15. The sum of the nagative rank numbers is —24. Table II shows that the probability of a sum of 24 or less is between 0.019 and 0.054 for 15 pairs. Fisher gives 0.0497 for the probability in this experiment by the t test.

The following data were obtained in a seed treatment experiment on wheat. The data are taken from a randomized block experiment with eight replications of treatments A and B. The figures in columns two and three represent the stand of wheat.

Block	*A*	*B*	*A-B*	*Rank*
1	209	151	58	8
2	200	168	32	7
3	177	147	30	6
4	169	164	5	1
5	159	166	—7	—3
6	169	163	6	2
7	187	176	11	5
8	198	188	10	4

The fourth column gives the differences and the fifth column the corresponding rank numbers. The sum of the negative rank numbers is —3. Table II shows that the total 3 indicates a probability between 0.024 and 0.055 that these treatments do not differ. Analysis of variance leads to a least significant difference of 14.2 between the means of two treatments for 19:1 odds, while the difference between the means of A and B was 17.9. Thus it appears that with only 8 pairs this method is capable of giving quite accurate information about the significance of differences of the means.

Discussion. The limitations and advantages of ranking methods have been discussed by Fried-

ıan (3), who has described a method for testng whether the means of several groups differ ignificantly by calculating a statistic χ^2_r from he rank totals. When there are only two groups to be compared, Friedman's method is equivalent to the binomial test of significance based on the number of positive and negative differences in a series of paired comparisons. Such a test has been shown to have an efficiency of 63 percent (1). The present method or comparing the means of two groups utilizes nformation about the magnitude of the differnces as well as the signs, and hence should ave higher efficiency, but its value is not nown to me.

The method of assigning rank numbers in he unpaired experiments requires little explaation. If there are eight replicates in each roup, rank numbers 1 to 16 are assigned to the xperimental results in order of magnitude and where tied values exist the mean rank value is sed.

TABLE I

'or Determining the Significance of Differences in Unpaired Experiments

'o. of replicates	Smaller rank total	Probability for this total or less
5	16	.016
5	18	.055
6	23	.0087
6	24	.015
6	26	.041
7	33	.0105
7	34	.017
7	36	.038
8	44	.0104
8	46	.021
8	49	.050
9	57	.0104
9	59	.019
9	63	.050
10	72	.0115
10	74	.0185
10	79	.052

In the case of the paired comparisons, rank umbers are assigned to the differences in order f magnitude neglecting signs, and then those ınk numbers which correspond to negative ifferences receive a negative sign. This is ecessary in order that negative differences ıall be represented by negative rank numbers, and also in order that the magnitude of the rank assigned shall correspond fairly well with the magnitude of the difference. It will be recalled that in working with paired differences, the null hypothesis is that we are dealing with a sample of positive and negative differences normally distributed about zero.

The method of calculating the probability of occurrence of any given rank total requires some explanation. In the case of the unpaired experiments, with rank numbers 1 to $2q$, the possible totals begin with the sum of the series 1 to q, that is, $q(p+1)/2$; and continue by steps of one up to the highest value possible, $q(3q+1)/2$. The first two and the last two of these totals can be obtained in only one way, but intermediate totals can be obtained in more than one way, and the number of ways in which each total can arise is given by the number of q-part partitions of T, the total in question, no part being repeated, and no part exceeding $2q$. These partitions are equinumerous with another set of partitions, namely the partitions of r, where r is the serial number of T in the possible series of totals beginning with 0, 1, 2, . . . r, and the number of parts of r, as well as the part magnitude, does not exceed q. The latter partitions can easily be enumerated from a table of partitions such as that given by Whitworth (5), and hence serve to enumerate the former. A numerical example may be given by way of illustration. Suppose we have 5 replications of measurements of two quantities, and rank numbers 1 to 10 are to be assigned to the data. The lowest possible rank total is 15. In how many ways can a total of 20 be obtained? In other words, how many unequal 5-part partitions of 20 are there, having no part greater than 10? Here 20 is the sixth in the possible series of totals; therefore $r = 5$ and the number of partitions required is equal to the total number of partitions of 5. The one to one correspondence is shown below:

Unequal 5-part partitions of 20	*Partitions of 5*
1-2-3-4-10	5
1-2-3-5-9	1-4
1-2-3-6-8	2-3
1-2-4-5-8	1-1-3
1-2-4-6-7	1-2-2
1-3-4-5-7	1-1-1-2
2-3-4-5-6	1-1-1-1-1

By taking advantage of this correspondence, the number of ways in which each total can be obtained may be calculated, and hence the probability of occurrence of any particular total or a lesser one.

The following formula gives the probability of occurence of any total or a lesser total by chance under the assumption that the group means are drawn from the same population:

$$P=2\left\{1+\sum_{i=1}^{i=r}\sum_{j=1}^{j=q}\sqcap_j^i-\sum_{n=1}^{n=r-q}\left[(r-q-n+1)\sqcap_{q-1}^{q-2+n}\right]\right\}\Big/\frac{\lfloor 2q}{\lfloor q\times\lfloor q}$$

$\sqcap_j^i$ represents the number of j-part partitions of i,
r is the serial number of possible rank totals, 0, 1, 2, $\cdots$ r.
q is the number of replicates, and
n is an integer representing the serial number of the term in the series.

In the case of the paired experiments, it is necessary to deal with the sum of rank numbers of one sign only, + or —, whichever is less, since with a given number of differences the rank total is determined when the sum of + or — ranks is specified. The lowest possible total for negative ranks is zero, which can happen in only one way, namely, when all the rank numbers are positive. The next possible total is —1, which also can happen in only one way, that is, when rank one receives a negative sign. As the total of negative ranks increases, there are more and more ways in which a given total can be formed. These ways for any totals such as —r, are given by the total number of unequal partitions of r. If r is 5, for example, such partitions, are 5, 1-4, 2-3. These partitions may be enumerated, in case they are not immediately apparent, by the aid of another relation among partitions, which may be stated as follows:

The number of unequal j-part partitions of r, with no part greater than i, is equal to the number of j-part partitions of $r-\binom{j}{2}$, parts equal or unequal, and no part greater than $i-j+1$ (4).

For example, if r equals 10, j equals 3, and i equals 7, we have the correspondence shown below:

Unequal 3-part partitions of 10	1-2-7 1-3-6 1-4-5 2-3-5
3-part partitions of 10—3, or 7, no part greater than 5	1-1-5 1-2-4 1-3-3 2-2-3

The formula for the probability of any given total r or a lesser total is:

$$P=2\left[1+\sum_{n}\left(\sum_{i=n}^{i=r-\binom{n}{2}}\sqcap_n^i\right)\right]\Big/2^q$$

r is the serial number of the total under consideration in the series of possible totals 0, 1, 2, . . ., r,

q is the number of paired differences.

In this way probability tables may be readily

TABLE II

For Determining the Significance of Differences in Paired Experiments

Number of Paired Comparisons	Sum of rank numbers, + or —, whichever is less	Probability of this total or less
7	0	0.016
7	2	0.047
8	0	0.0078
8	2	0.024
8	4	0.055
9	2	0.0092
9	3	0.019
9	6	0.054
10	3	0.0098
10	5	0.019
10	8	0.049
11	5	0.0093
11	7	0.018
11	11	0.053
12	7	0.0093
12	10	0.021
12	14	0.054
13	10	0.0105
13	13	0.021
13	17	0.050
14	13	0.0107
14	16	0.021
14	21	0.054
15	16	0.0103
15	19	0.019
15	25	0.054
16	19	0.0094
16	23	0.020
16	29	0.053

prepared for the 1 percent level or 5 percent level of significance or any other level desired.

Literature Cited

1. Cochran, W. G., The efficiencies of the binomial series tests of significance of a mean and of a correlation coefficient. Jour. Roy. Stat. Soc., 100:69-73, 1937.
2. Fisher, R. A. The design of experiments. Third ed., Oliver & Boyd, Ltd., London, 1942.
3. Friedman, Milton. The use ranks to avoid the assumption of normality. Jour. Amer. Stat. Assn. 32:675-701. 1937.
4. MacMahon, P. A. Combinatory analysis, Vol. II, Cambridge University Press. 1916.
5. Whitworth, W. A. Choice and chance, G. E. Stechert & Co., New York, 1942.

TEACHING AND RESEARCH AT THE STATISTICAL LABORATORY, UNIVERSITY OF CALIFORNIA

The Statistical Laboratory of the University of California, Berkeley, was established in 1939 as an agency of the Department of Mathematics. The functions of the Laboratory include its own research, help in the research carried on in other institutions, and a cycle of courses and of exercises for students.

The outbreak of the war soon after the establishment of the Statistical Laboratory influenced the direction of its research to a considerable extent. War problems were studied unofficially first and then under a contract with the National Defense Research Committee. These activities trained a considerable number of persons who were later absorbed by the Services. Also, the Laboratory acquired an efficient set of computing machines and other equipment.

In the future, the Laboratory's own research will be concerned with developing statistical techniques on one hand and with analyses of applied problems on the other.

Cooperation with other institutions is based on the principle of free choice and, therefore, care is taken to avoid anything suggesting a tendency to centralize statistical research in the Laboratory of the Department of Mathematics to the exclusion or the restraint of such research in other Departments of the University. The Statistical Laboratory has a hand in pieces of research for which the experimenter requests statistical help. The help offered consists primarily of advice. However, in cases where the numerical treatment of the problem is complex, computations are performed in the Laboratory.

Because of the voluntary basis of cooperation, contacts with institutions of experimental research are not systematic. The closest and most fruitful contacts thus far have been with the California Forest and Range Experiment Station located on the Berkeley Campus. Perhaps unexpectedly, in addition to practical results, these contacts originated a purely theoretical paper in the Annals of Mathematical Statistics. Other theoretical papers are expected because it appears that forest and range experimentation involves specific and difficult problems which require new statistical methods. Also, in contact with the Department of Entomology, some very interesting problems were found. The work done on them is also reflected in some publications.

The teaching in the Statistical Laboratory is geared to train research workers and teachers of statistics. The cycle of courses and of laboratory work now offered is essentially that planned in 1939, but some changes are under consideration. Both the original set-up and the reforms considered are based on the following premises.

The first and generally admitted premise is that a university teacher must be a research worker, that is to say, must be effectively capable of inventive work. As applied to statistics such inventive work may be of two kinds. First, inventiveness may express itself in developing new sections of the mathematical theory of statistics. In the present state of our science this requires not only the knowledge of the existing theory of statistics but also a considerable mastery of the theory of functions and other branches of mathematics. Next, the inventiveness may concern the techniques in statistical design of experiments or of observation in relation to the already existing statistical tools. Here the success of the research depends on a thorough knowledge of the tools and also of the particular domain in which they have to be used.

It is obvious that proficiency in the first of these items is especially desirable for a teacher

of statistics in a department of mathematics. A proficiency in the second, combined with a profound knowledge of the field of application, would be required for a teacher of statistics in a department of an experimental science. Both these directions of training are represented in the work of the Statistical Laboratory.

In order to train future research workers in mathematical statistics, in the strict sense of the word, several courses, both graduate and undergraduate, are offered. They are adjusted to courses in pure mathematics prerequisite to an intensive study of probability and statistics. A special course in the theory of functions was introduced in the Department of Mathematics to cover certain sections of the theory particularly important in studies of probability. Also, a certain number of hours of laboratory work are given to supplement the purely mathematical education of students.

In order to train future research workers in applied statistics who, due to their specialized training in experimental sciences, cannot be expected to be equally proficient in mathematics, special graduate applicational courses are offered. The purpose of these courses is to make the students understand all the "whys" and the "wherefores" of the statistical theories developed for them, without entering into complex mathematical proofs. To attain this aim special didactical methods are used. Theorems are always stated in exact terms. If the proof is easy, it is also given in an exact form. Otherwise a sampling experiment is performed in class to check the theorem and to create an intuitive feeling of the machinery behind it. Experience shows that, although somewhat slow, this method is very effective.

Instruction of this kind is given in relation to applied problems of particular interest to the majority of students in the class. Since the applied sciences are still a most fruitful source of theoretical statistical problems, the students of mathematical statistics are advised to attend applicational courses.

Studies in statistics are mostly on the graduate level and the University offers two higher degrees in statistics, M. A. and Ph. D.

Jerzy Neyman

QUERIES

QUERY: I have been told that the use of Yates' correction in χ^2 calculations involving frequency data for 1 d.f. always results in an overcorrection, sometimes serious, in cases where such χ^2's are added together. It would follow that the individual estimates of χ^2 obtained by this method are negatively biased and if this is true the corresponding P values must be positively biased because the direction of curvature of cumulative χ^2 plotted against P is such as to accentuate rather than diminish the bias. Queries: (1) Is the corrected estimate of P, for 1 d.f., always better than the corrected estimate, even in cases involving many categories as, for example, the 2x2x2 table discussed by Snedecor in the August Bulletin? (2) If a large number of uncorrected χ^2's each with 1 d.f. is summed, is the bias likely to build up so as to indicate highly significant deviations where none exists? (3) Is it possible that the subtraction of some number smaller than 0.5 from individual deviations would give, if not consistently smaller bias, perhaps generally better results?

ANSWER: (1) No, the estimate of P based on the corrected value is not always better, but is an improvement in the majority of those situations where the correction is suitable. In the supplement to the Journal of the Royal Statistical Society, Volume 1, pages 217-235 (1934), Yates considered two cases, the binomial and the 2x2 contingency table, each with a single degree of freedom. Of these he concluded that, ". . . no consistent under- or overestimate of significance will be made when applying the ordinary test corrected for continuity to a heterogeneous collection of data, even with very small expectations." For any particular set of such data, the effect of the correction can be determined from the exact tests described by Yates and also by Fisher and by Goulden in their texts.

(2) No, biases are ordinarily negligible if uncorrected values are added. On the other hand, Cochran (Iowa State College Journal of Science, Volume 16, pages 421-436, 1942) gave an illustration of the violent bias that could be introduced by adding corrected values of χ^2.

(3) For the cases discussed by Yates he remarks, "... it is clear that no simple modification of the correction for continuity will materially improve the approximations obtained." These are the cases that are most common if a single degree of freedom is involved. Cochran (loc. cit.) gives the general rule in applying corrections for continuity: "Calculate χ^2 by the usual formula. Find the next lowest possible value of χ^2 to the one to be tested, and use the tabular probability for a value of χ^2 midway between the two." A correction different from one-half is shown to be suitable in testing for linkage in a 9:3:3:1 segregation.

GEORGE W. SNEDECOR

QUERY: We have an experiment on flies laid out in 7 blocks of 4 plots each. The treatments were sprays containing, respectively, 4, 8 and 16 units of the active ingredient, designed to kill adult flies as they emerged from the breeding medium. The blocks comprised 7 sources of the medium. Numbers of adults found in cages set over the plots are as follows:

	Treatments			
Block	T_1	T_2	T_3	Check
1	445	414	247	423
2	113	127	147	326
3	122	206	138	246
4	227	78	148	141
5	132	172	356	208
6	31	45	29	303
7	177	103	63	256
Total	1247	1145	1128	1903

Analysis of Variance

	D.F.	S.S.	M.S.
Block	6	195851	
Treatment	3	58228	19409
Error	18	122110	6784

Casual inspection shows that there is little variation among the sums for the treated plots, but that the total for the untreated is somewhat greater. I have a feeling that the analysis of variance does not test this point.

(A) Is there a simple modification of the analysis of variance for testing the significance of the difference between treatment and check?

(B) Since the data are discrete, do you think a transformation (perhaps square root) would help?

ANSWER: (A) In an experiment of this design, the degrees of freedom for treatments may always be divided into two parts, a single degree for the comparison, treatment vs. check, and the remainder pertaining to differences among the treatment means (Fisher, section 42; Snedecor, Chapter 15). Assuming equal intervals of treatment effects, the mean square for the single degree of freedom may be calculated as follows:

$$3(1903) - (1247) + 1145 + 1128) = 2189$$

$$\frac{(2189)^2}{(3^2 + 1^2 + 1^2 + 1^2)\ (7)} = 57044$$

The sum of squares for treatments is now subdivided in this manner:

Treat. vs. ck.	1	57044	57044
Treatments	2	1184	592
Total	3	58228	

As you observed, the treatment means are remarkably similar, whereas the mean for the check plots is significantly greater than that for the treated.

(B) Since some of your counts are small, tests of significance might be slightly more accurate with the square root transformation. However, this is unimportant in comparison with the heterogeneity apparent in your data. The values of χ^2 for the 3 treatment counts in the several blocks are 61.52, 4.53, 25.63, 73.60, 129.75, 4.34 and 58.55. With 2 d.f. each, 5 of these are highly significant, the total χ^2 with 14 d.f. being 357.92. This might indicate treatment differences were it not for the inconsistency of the plot counts. The interaction or heterogeneity χ^2 for the 7 blocks of 3 treatments is 351.00, almost as large as the total above. This raises the question as to whether you have an adequate experimental technique. If, within the limits of error of the Poisson distribution, your counts are not repeatable under your experimental conditions, it would seem premature to attempt the evaluation of treatment effects.

GEORGE W. SNEDECOR

QUERY: I don't think any of us know just how to interpret correlation coefficients. A significant r seems to have little value unless one knows *exactly* how much correlation exists. The question revolves around the point as to what part of the total variation is accounted for by the correlation coefficient. One author says that an r of 0.5 accounts for 25% of the total variation, while another states that the same r would give 86% unaccounted for leaving 14% accounted for.

What we want is a more definite agreement as to the percentage due to regression. Why don't we do away with r and use something like Ezekiel's coefficient of determination?

ANSWER: The confusion pointed out by querist springs directly from use of the non-specific word, variation; but indirectly there is failure to distinguish clearly between sample and population, between correlation and regression, between exact and approximate statements about measures of variation. First, let us consider doing away with r.

The correlation coefficient, r, calculated from a random sample, is an estimate of ρ, a parameter of the normal, bivariate population; and ρ, in turn, is a measure of co-variation in the population. The tendency to vary together is often attributable to common causes, such as genes inherited by siblings. In certain cases, the correlation coefficient is the fraction of causes which are common to pairs of individuals.

Sample estimates, r, are subject to sampling variation, so that tests of hypotheses and fiducial statements are useful. Furthermore, several sample r's from the same population may be combined into a more reliable estimate than any one of them. Incidentally, one cannot learn from a sample *exactly* how much correlation exists.

The two foregoing paragraphs indicate that r has meaning and utility aside from its association with analysis of variance in regression. We are not likely to do away with r because of some confusions that have entangled it.

As for regression, it can be calculated and interpreted without the use of the correlation coefficient. If x and y denote deviations from the means, $\bar{x}$ and $\bar{y}$, of the variates X and Y, then $(\Sigma xy)^2/\Sigma x^2$ is the portion of Σy^2 associated with regression, the remainder being the sum of the squares of the deviations from regression.

It is an incident of the algebra that $r^2\Sigma y^2$ turns out to be equal to $(\Sigma xy)^2/\Sigma x^2$, and $(1 - r^2)\Sigma y^2$ to the remainder of the sum of squares. In this sense, then, r^2 is the fraction of the variation accounted for by regression, the term, "variation", being here synonymous with the *sum of squares* of the dependent variate, Y.

The analysis of the variance of Y is as follows:

Source	D.F.	S.S.	M.S.
Regression	1	$r^2\Sigma y^2$	$r^2\Sigma y^2$
Deviations	n—2	$(1-r^2)\Sigma y^2$	$\frac{(1-r^2)\Sigma y^2}{n-2}$
Total	n—1	Σy^2	$\frac{\Sigma y^2}{n-1}$

From this it is clear that the mean squares for regression and deviations do not add to the total mean square. That is, mean square is a measure of variation that is not divided into two parts that add up.

The *variance* in regression can be partitioned into two additive parts by this device:

$$\frac{\Sigma y^2}{n-1} - \frac{(1-r^2)\Sigma y^2}{n-2} = s_y^2\left\{1 - \frac{n-1}{n-2}(1-r^2)\right\},$$

where s_y^2 is put equal to the variance, $\frac{\Sigma_y^2}{n-1}$. If an *adjusted* value of the correlation coefficient, r_A^2, is substituted for $1 - \frac{n-1}{n-2}(1-r^2)$, then it may be said that r_A^2 is the fraction of s_y^2 due to regression, the remaining fraction, $(1-r_A^2)$ being associated with deviations from regression. This is a second sense in which the correlation coefficient is related to the fraction of the total variation due to regression. Now it is r_A^2 instead of r^2 that is the correct fraction, and "variation" means *variance*.

In large samples the fraction, $\frac{n-1}{n-2}$, is approximately 1, r_A^2 is approximately equal to r^2, and s_y^2 is divided approximately into the

two portions, $r^2s_y^2$ and $(1-r^2)s_y^2$. In this third sense r^2 is the fraction of the total variation associated with regression, "variation" again being *variance;* but here the relation is only approximate.

Finally, the standard deviation can be multiplied by r and $\sqrt{1-r^2}$, and these quantities may be said to be fractions of the total variation if by "*variation*" is meant *standard deviation.* These fractions aren't very useful, chiefly because they do not add to unity. Again, the relations are approximate from the viewpoint of estimation.

In conclusion: (i) In making precise statements it is necessary to specify what measure of variation is involved. (Note: I have not considered the measure, $s_y^2=\Sigma y^2/n$.) (ii) If you are interested in correlation, don't confuse the issue by introducing ideas related to variance in regression. (iii) If it is the prediction of one variate from another that is interesting, stick to regression methods—these are quite independent of correlation.

George W. Snedecor

A LETTER TO THE MEMBERS OF THE BIOMETRICS SECTION

This issue of the BIOMETRICS BULLETIN is being mailed to nearly 1000 members and to more than 100 subscribers. We could have no better proof of the active interest in statistical methods among research workers in the biological fields. Our members also include many who are not biologists but have joined because of the importance of the statistical methods developed initially for solving biological problems. This large support in the first year of the BIOMETRICS BULLETIN augurs well for its future. Equally, it emphasizes the task faced by the Editorial Board in meeting the needs of so large and varied a membership.

In the past the activities of the Biometrics Section were restricted to arranging programs at meetings of the parent Association and of various biological societies. Little organization was needed beyond a Chairman, a Secretary and a few Committee Members. With the launching of the BIOMETRICS BULLETIN, however, the Section has outgrown this organizational pattern. A special committee has been appointed under the chairmanship of Dr. A. E. Brandt, Soil Conservation Service, Washington, D. C., to draft a constitution, so that we can participate effectively in the larger plans now being developed for the parent Association. This will be the main item on the agenda of the business meeting at Cleveland, announced elsewhere in this issue.

The first year of the BIOMETRICS BULLETIN has been a difficult one. The Section is indebted especially to the able direction of our Editor, Professor Gertrude M. Cox, to our Secretary, Dr. Horace W. Norton, the Secretary's office of the American Statistical Association and the members of the Editorial Board and others whose collaboration has made the first year possible. As the BIOMETRICS BULLETIN becomes better known, material will come more easily than during the first year. With the resumption of meetings, abstracts of the papers on the programs will be published, together with some papers in full. In the near future, the A. S. A. Bulletin will be sent to associate as well as to regular members of the Section. As opportunity arises, some features now appearing in the BIOMETRICS BULLETIN will be shifted to the A. S. A. Bulletin, releasing further space for papers on biometrics. We depend upon the active cooperation of our members to make the BIOMETRICS BULLETIN an influential and valuable source of biometrical methods.

The fields of interest represented in the Section have been tabulated from the information on our membership forms. Where a member checked several fields of interest, each has been assigned a fractional part in determining the number of members in each field. No information was available for 210 of the 979 regular and associate members. The following table indicates in percentages the interests of the remaining 769 members.

1. Biology		65.8
General biology	15.3	
Experimental biology	8.4	
Microbiology	0.7	
Forestry	0.8	
Economic biology	2.2	
Agricultural biology	12.4	

Medical sciences	26.0	
2. Meteorology		1.6
3. Chemistry		4.4
4. Mathematics		4.5
5. Industry		5.6
6. Social Sciences		18.1

Several biological fields in which statistical techniques are important have as yet few representatives in the Section. Many new members could be added from these groups. We hope that nonbiological members will find enough of interest in the BIOMETRICS BULLETIN to justify continuing their memberships; certainly the biometricians stand to gain from their cooperation.

With the continued active cooperation of biologists and statisticians the BIOMETRICS BULLETIN can aid materially in improving the quality and efficiency of biological work through statistical methods.

C. I. Bliss, *Chairman*
November 23, 1945. Biometrics Section

MEETINGS OF THE BIOMETRICS SECTION

With the end of the war, scientific organizations are resuming their meetings although they are not yet on their standard schedules. The Biometrics Section plans to participate in three meetings in the first quarter of 1946.

The first is that of the American Statistical Association, which is the deferred Annual Meeting for 1945. It will be held at Cleveland, Ohio on January 24 to 27 with the Statler Hotel as headquarters. Two half-day sessions sponsored by the Biometrics Section are scheduled for Friday, January 25. One will concern problems of biological assay and the other, to be held jointly with the Institute of Mathematical Statistics, will stress the analysis of certain incomplete experimental designs. Of particular importance is the luncheon and business meeting which will be held also on Friday to discuss a constitution for the Section and to elect officers for 1946. We hope that all Section members attending the Cleveland meetings will participate in this business session.

The second meeting is that of the Federation of American Societies for Experimental Biology which has been scheduled for Atlantic City during the week of March 11. These arrangements have been made so recently that there has not been time to plan a joint program with one of the Federation Societies. So many of our members are interested in pharmacological and other phases of medical research that a session of biometrical interest would have a wide appeal.

The third meeting is that of the American Association for the Advancement of Science in St. Louis on March 27 to 30. A special program committee has been appointed for the Biometrics Section with Dr. H. C. Fryer, Kansas State College, Manhattan, as Chairman. We hope to have a joint session with the Ecological Society of America and with other participating organizations, as well as several sessions of our own.

All of these meetings have been scheduled on much shorter notice than usual. We appeal to members of the Section both for papers and for suggestions as to possible speakers and topics. Those for the St. Louis meetings should be sent directly to Dr. Fryer and for the others to the Chairman or the Secretary of the Section. If time limitations prevent our scheduling as many papers as may become available, those which otherwise would be suitable will be read by title and their abstracts printed in the BIOMETRICS BULLETIN.

NEWS AND NOTES

A statistical summer session is to be held at the Institute of Statistics, Raleigh, N. C., June 17 to July 26, 1946. R. A. FISHER, C. I. BLISS, W. G. COCHRAN, CERTRUDE COX, G. W. SNEDECOR and J. WOLFOWITZ will offer courses . . . After completing his Ph.D. work at Iowa State College, JEROME C. R. LI went to Queens College of the City of New York where he is an instructor in mathematics. In addition to mathematics courses, he teaches one course in mathematical statistics and one in probability . . . CWO D. D.

MASON was released from the army where he was a soil physicist with the Engineer Corps. He has resumed the study of meteorological-soil-plant relations with particular emphasis on the effect of meteorological factors on soil moisture. Mr. Mason is working towards a Ph.D. degree in soil physics at North Carolina State College . . . R. E. BLASER, Florida Agricultural Experiment Station is working at the same place on his Ph.D. in Agronomy. . . WILLIAM P. MARTIN left the University of Arizona in February 1945 to go to the U. S. Regional Salinity Laboratory at Riverside, California as Associate Microbiologist. Even wonderful California and associations with CECIL H. WADLEIGH, OSCAR C. MAGISTAD and E. R. PARKER could not keep Mr. Martin away from Arizona. He is now with the Southwestern Forest and Range Experiment Station, at Tucson, in charge of the Forest and Range Influences Section. He says, "The work is exceedingly interesting. We have one very fine substation for watershed management research in the Sierra Ancha Experimental Forest. Our installations of ten years duration range from small lysimeters up to paired intermittent- and perennial-stream watersheds of several hundred acres. Elevations range from 2200 up to 7700 feet and vegetation zones from desert shrub up to pine-fir forest" . . . CHURCHILL EISENHART, (who has promised us an article for the Bulletin soon) statistician of the Wisconsin Agricultural Experiment Station, has returned to his post after almost three years during which he was engaged in war research at Tufts College and at Columbia University. At Columbia he was a member first of the Applied Mathematics Group, later of the Statistical Research Group (SRG) . . . KENNETH J. ARNOLD, formerly of the Mathematics Department of the University of New Hampshire and later of SRG-Columbia has accepted a position as assistant professor in the Department of Mathematics at the University of Wisconsin. . . J. WOLFOWITZ who was also with SRG-Columbia has joined the staff of the Institute of Statistics, University of North Carolina . . . R. L. ANDERSON, recently with SRG-Princeton, has his voice resounding in Patterson Hall, new residence of the Institute of Statistics . . . ALBERT E. WAUGH, head of the Department of Economics and Professor of Statistics at the University of Connecticut, has been appointed Dean of the College of Arts and Sciences at that institution . . . MAJOR EUGENE L. HAMILTON left his position as Chief of the Vital Statistics Division in the Department of Public Health, Dallas, Texas on July 1942 to enter the army. He has recently returned from overseas where he spent twenty months as Medical Statistician for General Courtney H. Hodges' First Army. He is now on duty in the Office of the Surgeon General, Washington, D. C. While in England, before embarking to take part in the assault landings in Normandy, Major Hamilton was elected a Fellow of the Royal Statistical Society . . . From a reliable source we have been informed "Just now BOYD HARSHBARGER is out in Shenandoah County making a nuisance of himself among the farmers, finding out how much money they are making, what threats they find most effective in getting their socks darned, and various other items of importance in the planning of the post-war world." Mr. Harshbarger, statistician at Virginia Polytechnic Institute reports that D. B. REID from the Canadian Research Council joined their group in October as a graduate assistant in Statistics. Mr. Reid was acting statistician for the Council in the absence of J. W. HOPKINGS who was on leave to England. Mr. Harshbarger and D. B. DeLURY are giving three classes and conducting a seminar in mathematical statistics at VPI. In January they are sponsoring a four-week conference in statistical method for the Experiment Station staff. Mr. DeLury writes, "At the moment, a number of circumstances keep me tethered on a pretty short leash, but one of these days, like the old monk in Siberia, I'm going to break lose with a yell." . . . A. E. BRANDT returned to the Soil Conservation Service, Department of Agriculture on September 24 . . . LT. H. F. (COTTON) ROBINSON was instructor of navigation at Miami Beach until released from the Navy. He has joined the staff of the Institute of Statistics as Assistant Professor in the Division of Plant Science Statistics, which is headed by J. A. RIGNEY . . . The meeting of the Editorial Committee of this Bulletin, which was held in Washington, Hotel Statler, October 13 was closed with the poem

"Said Chester to Lester" . . . BESSE B. DAY, formerly of the Forest Service, engaged in war research with the Applied Physics Laboratory, John Hopkins University. Miss Day was a statistical consultant on the VT-Fuze project. This was a radio-controlled fuze used to detonate artillery shells. It was particularly effective against the V-1 bombs over London. The anti-aircraft effectiveness of VT-fuzed projectiles was more than 300 percent greater than that of time-fuzed ones. The rugged and very small vacuum tube and battery which were developed to withstand the strain of being fired from a gun should have extensive post-war uses . . . About 250 letters were sent to those who gave no information regarding fields of interest when they subscribed to the Bulletin. The responses have been helpful and interesting. One person saw no connection between industrial and biological applications of statistical tools. Another person stated, "Candidly I think I should not subscribe to the Biometrics Bulletin." To these persons we might call attention to the fact that many men trained in biological statistics are doing industrial statistical work. This idea is expressed by J. R. CRAWFORD, Lockheed, Burbank, California. "The Bulletin contains unique statistical applications which may often be adapted to industrial use." We hope this statement holds true about our future issues . . . COLONEL JOHN G. BOOTON states, "Those of us who are interested in statistics as a tool of warfare try to follow developments of statistical methods for other purposes in order to adapt such as may be suitable to our problems" . . . CAPTAIN WALTER L. DEEMER writes, "The Bulletin is filling an important need. We have already used the query section and been greatly helped by Professor Snedecor's reply" . . . H. O. HETZER, Associate Animal Husbandman in Swine Investigations, Bureau Animal Industry, is doing work which centers primarily on research in swine breeding and genetics. Recently, he has been working on the inheritance of swine colors . . . ROY A. BAIR, Associate Agronomist, Everglades Experiment Station, Belle Glade, Florida, took a trip last summer to visit the Experiment Stations of North and South Carolina, Georgia, Alabama, the Federal Station at Tifton, Georgia and the Coker Seed Co. at Hartsville, S. C. The objectives of this trip were (1) to find improvements for small grain harvesting and threshing equipment adapted to small plot work, (2) to locate disease-resistant varieties of barley; and water tolerant pasture grasses, and (3) to find an agronomist ready to stop existing somewhere else and start living in Florida. Tabulated results of his trip are (1) yes, (2) yes, and (3) no! . . . D. J. FINNEY, recently moved to 7 Kebel Road, Oxford, England, writes "the arrangements for my removal have kept me very fully occupied for the last three months. Even now, though I have arrived here, I am far from settled, as I am without heat or electricity in these rooms, have no typist or computing assistance, and only a minimum of furniture."

Officers of the American Statistical Association, President: Walter A. Shewhart, Directors: Henry B. Arthur, C. I. Bliss, Simon Kuznets, E. Grosvenor Plowman, Willard L. Thorp, and Helen M. Walker; Vice-Presidents, William G. Cochran, A. D. H. Kaplan, Lowell J. Reed; Secretary-Treasurer, Lester S. Kellogg.

Officers of the Biometrics Section: C. I. Bliss, Chairman; H. W. Norton, Secretary.

Editorial Committee for the BIOMETRICS BULLETIN: Gertrude M. Cox, Chairman; C. I. Bliss, W. G. Cochran, F. R. Immer, J. Neyman, H. W. Norton, L. J. Reed, G. W. Snedecor, Sewall Wright.

Material for the BULLETIN should be addressed to the Chairman of the Editorial Committee, Institute of Statistics, North Carolina State College, Raleigh, N. C., material for Queries should go to "Queries", Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.

BIOMETRICS BULLETIN

Contents of Volume I

INDEX